

Representation formula for the entropy and functional inequalities

Joseph Lehec *

July 18, 2011

Abstract

We prove a stochastic formula for the Gaussian relative entropy in the spirit of Borell's formula for the Laplace transform. As an application, we give simple proofs of a number of functional inequalities.

1 Introduction: Borell's formula

Let γ_d be the standard Gaussian measure on \mathbb{R}^d :

$$\gamma_d(dx) = \frac{e^{-|x|^2/2}}{(2\pi)^{d/2}} dx$$

where $|x| = \sqrt{x \cdot x}$ denotes the Euclidean norm of x . In [5, 6] Borell proves the following representation formula. Given a standard d -dimensional Brownian motion B and a bounded function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ we have

$$\log\left(\int_{\mathbb{R}^d} e^f d\gamma_d\right) = \sup_u \left[\mathbb{E}\left(f(B_1 + \int_0^1 u_s ds\right) - \frac{1}{2} \int_0^1 |u_s|^2 ds\right) \right], \quad (1)$$

where the supremum is taken over all random processes u , say bounded and adapted to the Brownian filtration. Among other applications, he derives easily the Prékopa-Leindler inequality. The name *Borell's formula* may be unfair to Boué and Dupuis who in an earlier paper [7] obtained a stronger result, allowing the function f to depend on the whole path $(B_t)_{t \in [0,1]}$ (see Theorem 9 below for a precise statement). Anyway, Borell and Boué-Dupuis agree that representation formulas such as (1) arose much earlier in optimal control theory, particularly in Fleming and Soner's work [14], and Borell should definitely be credited for bringing these techniques in the context of functional inequalities.

The present article deals with relative entropy. Let (Ω, \mathcal{A}, m) be a measured space and μ be a probability measure. The relative entropy of μ is defined by

$$H(\mu \mid m) = \int_{\Omega} \frac{d\mu}{dm} \log\left(\frac{d\mu}{dm}\right) dm \quad \text{if } \mu \ll m$$

*CEREMADE (UMR CNRS 7534) Université Paris-Dauphine.

and $H(\mu \mid m) = +\infty$ otherwise. It is well known that there is a Legendre duality between relative entropy and logarithmic Laplace transform:

$$H(\mu \mid m) = \sup_f \left(\int f \, d\mu - \log \left(\int_{\Omega} e^f \, dm \right) \right). \quad (2)$$

The purpose of this article is to prove a representation formula for the Gaussian relative entropy, both in \mathbb{R}^d and in the Wiener space, providing the entropy counterparts of the results mentioned above. All these formulas have a common feature: Girsanov's theorem. However, our approach is somewhat different from that of Borell and Boué-Dupuis: it draws a connection with the work of Föllmer [15, 16] which makes the whole argument arguably simpler. As an application, we give new, unified and simple proofs of a number of Gaussian inequalities.

2 Representation formula for the entropy

This section contains the main results of the article. Let us recall a couple of classical facts about relative entropy, see for instance [24, section 10] and the references therein. If \mathcal{A} is the Borel σ -field of a Polish topology on Ω then it is enough to take the supremum over bounded and continuous function in (2). In particular the map $\mu \mapsto H(\mu \mid m)$ is lower semicontinuous with respect to the topology of weak convergence of measures. If $T: (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ is a measurable map then

$$H(\mu \circ T^{-1} \mid m \circ T^{-1}) \leq H(\mu \mid m) \quad (3)$$

and assuming that $H(\mu \mid m) < +\infty$, equality occurs if and only if the density $d\mu/dm$ is a function of T .

We now describe the setting of the article. Let \mathbb{W} be the space of continuous paths

$$\{w \in \mathcal{C}^0(\mathbb{R}_+, \mathbb{R}^d), w_0 = 0\}$$

equipped with the topology of uniform convergence on compact intervals. Let \mathcal{B} be the associated Borel σ -field and let γ be the Wiener measure on $(\mathbb{W}, \mathcal{B})$. Let $x_t: w \mapsto w_t$ be the coordinate process and $(\mathcal{G}_t)_{t \geq 0}$ be the natural filtration of x . It is well known that \mathcal{B} coincides with the smallest σ -field containing $\cup_{t \geq 0} \mathcal{G}_t$. Let \mathbb{H} be the Cameron-Martin space: a path U belongs to \mathbb{H} if there exists $u \in L^2([0, +\infty); \mathbb{R}^d)$ such that

$$U_t = \int_0^t u_s \, ds, \quad t \geq 0.$$

The norm of U in \mathbb{H} is then defined by

$$\|U\| = \left(\int_0^{+\infty} |u_s|^2 \, ds \right)^{1/2}.$$

The Cauchy-Schwarz inequality shows that the Hilbert space \mathbb{H} embeds continuously in \mathbb{W} . Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)_{t \geq 0}$ we call *drift* any adapted process U which belongs to \mathbb{H} almost surely. Lastly, our Brownian motions are always d -dimensional, standard and always start from 0.

2.1 The upper bound

We shall use repeatedly Girsanov's formula, see [19, chapter 6].

Proposition 1. *Let B be a Brownian motion defined on some filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$ and let U be a drift. Letting μ be the law of $B + U$, we have*

$$H(\mu \mid \gamma) \leq \frac{1}{2} \mathbb{E} \|U\|^2. \quad (4)$$

Proof. Write $U_t = \int_0^t u_s \, ds$ and assume for the moment that $\|U\|^2 = \int_0^\infty |u_s|^2 \, ds$ is uniformly bounded. Then by Novikov's criterion

$$M_t = \exp\left(-\int_0^t u_s \cdot dB_s - \frac{1}{2} \int_0^t |u_s|^2 \, ds\right), \quad t \geq 0$$

is a uniformly integrable martingale and Girsanov's formula applies. Under

$$d\mathbb{Q} = M_\infty \, d\mathbb{P}$$

the process $X := B + U$ is a Brownian motion. Therefore X has law μ and γ under \mathbb{P} and \mathbb{Q} , respectively. Then by (3)

$$H(\mu \mid \gamma) \leq H(\mathbb{P} \mid \mathbb{Q}) = -\mathbb{E} \log(M_\infty) = \frac{1}{2} \mathbb{E} \|U\|^2,$$

which concludes the proof when $\|U\|$ is bounded. In the general case, define the stopping time

$$T_n = \inf(t \geq 0, \int_0^t |u_s|^2 \, ds \geq n),$$

let U_n be the stopped process $(U_n)_t = U_{t \wedge T_n}$ and μ_n be the law of $B + U_n$. With probability 1 we have $\|U\|^2 < +\infty$, thus $T_n \rightarrow +\infty$ and $U_n \rightarrow U$ in \mathbb{H} , hence in \mathbb{W} . Therefore $\mu_n \rightarrow \mu$ weakly. Also $\mathbb{E} \|U_n\|^2 \rightarrow \mathbb{E} \|U\|^2$ by monotone convergence. Thus, using the lower semicontinuity of the entropy (observe that \mathbb{W} is a Polish space)

$$\begin{aligned} H(\mu \mid \gamma) &\leq \liminf_n H(\mu_n \mid \gamma) \\ &\leq \liminf_n \frac{1}{2} \mathbb{E} \|U_n\|^2 = \frac{1}{2} \mathbb{E} \|U\|^2. \quad \square \end{aligned}$$

Remark. It follows immediately that when $\mathbb{E} \|U\|^2 < +\infty$, the law of $B + U$ is absolutely continuous with respect to the Wiener measure γ . Let us point out that this is actually true for all drifts U , even if $\mathbb{E} \|U\|^2 = +\infty$, see [19, chapter 7].

2.2 Föllmer's drift

Let us address the question whether, given a probability measure μ on \mathbb{W} , equality can be achieved in (4). Recall that $(x_t)_{t \geq 0}$ is the coordinate process on Wiener space $(\mathbb{W}, \mathcal{B}, \gamma)$ and that $(\mathcal{G}_t)_{t \geq 0}$ is its natural filtration. The following is due to Föllmer [15, 16].

Theorem 2. *Let μ be a measure on $(\mathbb{W}, \mathcal{B})$ having density F with respect to γ . There exists an adapted process u such that under μ the following holds.*

1. *The process $U_t = \int_0^t u_s \, ds$ belongs to \mathbb{H} almost surely.*
2. *The process $y = x - U$ is a Brownian motion.*
3. *The relative entropy of μ is*

$$H(\mu \mid \gamma) = \frac{1}{2} \mathbb{E}^\mu \|U\|^2.$$

We sketch the proof for completeness.

Proof. Throughout \mathbb{E}^γ and \mathbb{E}^μ denote expectations with respect to γ and μ respectively. On \mathcal{G}_t the measure μ has density

$$F_t := \mathbb{E}^\gamma(F \mid \mathcal{G}_t),$$

with respect to γ . A standard martingale argument shows that

$$\mu\left(\inf_{t \geq 0} F_t > 0\right) = \mu(F > 0) = 1. \quad (5)$$

Since Brownian martingales can be represented as stochastic integrals there exists an adapted process v satisfying

$$\gamma\left(\int_0^{+\infty} |v_s|^2 \, ds < +\infty\right) = 1 \quad (6)$$

and

$$F_t = 1 + \int_0^t v_s \cdot dx_s, \quad t \geq 0.$$

Let u be the process defined by

$$u_t = \mathbf{1}_{\{F_t > 0\}} (F_t)^{-1} v_t.$$

It is adapted and (5) and (6) yield

$$\mu\left(\int_0^\infty |u_s|^2 \, ds < +\infty\right) = 1,$$

which is the first assertion of the theorem.

The assertion 2 follows from Girsanov's formula, see [19, Theorem 6.2].

Under μ , we have

$$\begin{aligned} F_t &= 1 + \int_0^t F_s u_s \cdot dx_s \\ &= 1 + \int_0^t F_s u_s \cdot dy_s + \int_0^t F_s |u_s|^2 ds. \end{aligned}$$

Applying Itô's formula (recall that F is positive and y is a Brownian motion under μ) we obtain

$$\log(F) = \int_0^{+\infty} u_s \cdot dy_s + \frac{1}{2} \int_0^{+\infty} |u_s|^2 ds.$$

If $\mathbf{E}^\mu \|U\|^2 < +\infty$ the local martingale part in the equation above is integrable and has mean 0 so that

$$\mathbf{H}(\mu \mid \gamma) = \mathbf{E}^\mu \log(F) = \frac{1}{2} \mathbf{E}^\mu \|U\|^2.$$

Again, a localization argument shows that this equality remains valid when $\mathbf{E}^\mu \|U\|^2 = +\infty$, see [15, Lemma (2.6)]. \square

To finish this subsection, we give a formula for Föllmer's drift when the underlying density has a Malliavin derivative, we refer to the first chapter of [20] for the (little amount of) Malliavin calculus we shall use. For suitable $F: \mathbb{W} \rightarrow \mathbb{R}$ we let $DF: \mathbb{W} \rightarrow \mathbb{H}$ be the Malliavin derivative of F . The domain of D in the space $L^2(\mathbb{W}, \mathcal{B}, \gamma)$ is denoted by \mathbb{D}^2 . If $F \in \mathbb{D}^2$ then the Clark-Ocone formula asserts

$$\mathbf{E}^\gamma(F \mid \mathcal{G}_t) = 1 + \int_0^t \mathbf{E}^\gamma(D_s F \mid \mathcal{G}_s) \cdot dx_s, \quad t \geq 0.$$

We obtain the following result.

Lemma 3. *When $F \in \mathbb{D}^2$ the process u_t given by Theorem 2 is*

$$u_t = \frac{\mathbf{E}^\gamma(D_t F \mid \mathcal{G}_t)}{\mathbf{E}^\gamma(F \mid \mathcal{G}_t)} \mathbf{1}_{\{\mathbf{E}^\gamma(F \mid \mathcal{G}_t) > 0\}}.$$

This implies that μ -almost surely

$$u_t = \mathbf{E}^\mu\left(\frac{D_t F}{F} \mid \mathcal{G}_t\right).$$

2.3 Optimal drift in a strong sense

According to Theorem 2, the filtered probability space $(\mathbb{W}, \mathcal{B}, \mu, \mathcal{G})$ carries a Brownian motion y . The process $x = y + U$ has law μ and the drift U satisfies

$$\mathbf{H}(\mu \mid \gamma) = \frac{1}{2} \mathbf{E}^\mu \|U\|^2.$$

Still, it remains open whether *given* a probability space, a filtration and a Brownian motion, there exists a drift achieving equality in (4).

In this section, we show that this is indeed the case, under some restriction on the measure μ . The approach is taken from the article [4] in which Baudoin treats the case of Brownian bridges (see subsection 2.5 below). We refer to [21] for the background on stochastic differential equations.

Theorem 4. *Let B be a Brownian motion defined on some filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$. Let μ be a measure on \mathbb{W} , absolutely continuous with respect to γ and let $u_t: \mathbb{W} \rightarrow \mathbb{R}^d$ be the associated Föllmer process. If the stochastic differential equation*

$$X_t = B_t + \int_0^t u_s(X) \, ds, \quad t \geq 0 \quad (7)$$

has the pathwise uniqueness property, then it has a unique strong solution. This solution X satisfies the following.

1. *The process $U_t = \int_0^t u_s(X) \, ds$ belongs to \mathbb{H} almost surely.*
2. *The process X has law μ .*
3. *The relative entropy of μ is given by*

$$H(\mu \mid \gamma) = \frac{1}{2} \mathbb{E} \|U\|^2.$$

Proof. According to Theorem 2, on $(\mathbb{W}, \mathcal{B}, \mu)$ the coordinate process x satisfies

$$x_t = y_t + \int_0^t u_s(x) \, ds$$

where y is a Brownian motion. Therefore (7) has a weak solution. By Yamada and Watanabe's theorem, if pathwise uniqueness holds then (7) has a unique strong solution. Moreover, since pathwise uniqueness implies uniqueness in law, the solution X has law μ . The rest of Theorem 4 concerns the law of X , so it is contained in Theorem 2. \square

We end this section by showing that for a reasonably large class of measures μ , the stochastic differential equation (7) does satisfy the pathwise uniqueness property.

Definition 5. Let \mathcal{S} be the class of probability measures on $(\mathbb{W}, \mathcal{B}, \gamma)$ having a density of the form

$$F(w) = \Phi(w_{t_1}, \dots, w_{t_n}) \quad (8)$$

for some integer n , for some sample $0 \leq t_1 < t_2 < \dots < t_n$ and for some function $\Phi: (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ satisfying

- Φ is Lipschitz.

- $\nabla\Phi$ is Lipschitz.
- There exists $\epsilon > 0$ such that $\Phi \geq \epsilon$.

Lemma 6. *If μ belongs to \mathcal{S} then the equation (7) has the pathwise uniqueness property.*

Proof. Let μ have density F given by (8). Then $F \in \mathbb{D}^2$ and

$$DF(w) = \sum_{i=1}^n \nabla_i \Phi(w_{t_1}, \dots, w_{t_n}) \mathbf{1}_{[0, t_i]}$$

where $\nabla_i \Phi$ is the gradient of Φ in the i -th variable. By Lemma 3, the process associated to μ is

$$\begin{aligned} u_t(w) &= \frac{\mathbb{E}^\gamma(D_t F(w) \mid \mathcal{G}_t)}{\mathbb{E}^\gamma(F(w) \mid \mathcal{G}_t)} \\ &= \sum_{i=1}^n \frac{\mathbb{E}^\gamma(\nabla_i \Phi(w_{t_1}, \dots, w_{t_n}) \mid \mathcal{G}_t)}{\mathbb{E}^\gamma(\Phi(w_{t_1}, \dots, w_{t_n}) \mid \mathcal{G}_t)} \mathbf{1}_{[0, t_i]}(t). \end{aligned}$$

It is enough to prove that there is a constant C such that

$$|u_t(w) - u_t(\tilde{w})| \leq C \sup_{0 \leq s \leq t} |w_s - \tilde{w}_s|. \quad (9)$$

for all $t \geq 0$ and for all $w, \tilde{w} \in \mathbb{W}$. Fix $t \geq 0$ and assume that $t_k \leq t < t_{k+1}$ for some $k \in \{0, \dots, n-1\}$. By the Markov property of the Brownian motion

$$\mathbb{E}(\Phi(w_{t_1}, \dots, w_{t_n}) \mid \mathcal{G}_t) = \Psi(w_{t_1}, \dots, w_{t_k}, w_t)$$

where $\Psi(x_1, \dots, x_k, x)$ equals

$$\int_{\mathbb{W}} \Phi(x_1, \dots, x_k, x + w_{t_{k+1}-t}, \dots, x + w_{t_n-t}) \gamma(dw).$$

Then observe that $\|\Psi\|_{\text{lip}} \leq \|\Phi\|_{\text{lip}}$. We have a similar property when $0 \leq t < t_1$ and when $t_n \leq t$. The argument applies also to $\nabla_i \Phi$. The inequality (9) follows easily. \square

To sum up, we have the following representation formula.

Theorem 7. *Let $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$ be a filtered probability space and let $B: \Omega \rightarrow \mathbb{W}$ be a Brownian motion. For all $\mu \in \mathcal{S}$ we have*

$$H(\mu \mid \gamma) = \min_U \left(\frac{1}{2} \mathbb{E} \|U\|^2 \right)$$

where the minimum is on all drifts U such that $B + U$ has law μ .

2.4 The Boué and Dupuis formula

In this subsection the previous results are translated in terms of log-Laplace using the following lemma.

Lemma 8. *Let $f: \mathbb{W} \rightarrow \mathbb{R}$ bounded from below. For every positive ϵ there exists $\mu \in \mathcal{S}$ such that*

$$\log\left(\int_{\mathbb{W}} e^f d\gamma\right) \leq \int_{\mathbb{W}} f d\mu - H(\mu \mid \gamma) + \epsilon. \quad (10)$$

Proof. By monotone convergence we can assume that f is also bounded from above, and that $\int e^f d\gamma = 1$. Set $F = e^f$ and let μ be a probability measure on \mathbb{W} . Using $t \log(t) \leq |t - 1| + |t - 1|^2/2$ we get

$$\begin{aligned} H(\mu \mid \gamma) - \int_{\mathbb{W}} f d\mu &\leq \int_{\mathbb{W}} \left| \frac{G}{F} - 1 \right| F d\gamma + \frac{1}{2} \int_{\mathbb{W}} \left| \frac{G}{F} - 1 \right|^2 F d\gamma \\ &\leq \|F - G\|_{L^1(\gamma)} + C \|F - G\|_{L^2(\gamma)}^2 \end{aligned}$$

where G is the density of μ and C is some constant (recall that f is bounded below). Therefore, it is enough to prove that there exists $\mu \in \mathcal{S}$ whose density G is arbitrarily close to F in $L^2(\gamma)$. This is left to the reader. \square

Here is the Boué and Dupuis formula.

Theorem 9. *For every function $f: \mathbb{W} \rightarrow \mathbb{R}$ measurable and bounded from below, we have*

$$\log\left(\int_{\mathbb{W}} e^f d\gamma\right) = \sup_U \left[\mathbb{E}\left(f(B + U) - \frac{1}{2}\|U\|^2\right) \right],$$

where the supremum is taken over all drifts U .

This is actually slightly more general than the result in [7], which concerns the space $\mathcal{C}([0, T], \mathbb{R}^d)$ for some finite time horizon T .

Proof. Let U be a drift and μ be the law of $B + U$. By Proposition 1 and the entropy/log-Laplace duality

$$\mathbb{E}\left(f(B + U) - \frac{1}{2}\|U\|^2\right) \leq \int_{\mathbb{W}} f d\mu - H(\mu \mid \gamma) \leq \log\left(\int_{\mathbb{W}} e^f d\gamma\right).$$

On the other hand, given $\epsilon > 0$, there exists a probability measure $\mu \in \mathcal{S}$ satisfying (10). Since $\mu \in \mathcal{S}$, Theorem 7 asserts that there exists a drift U such that $B + U$ has law μ and satisfying

$$H(\mu \mid \gamma) = \frac{1}{2} \mathbb{E}\|U\|^2.$$

Then (10) becomes

$$\log\left(\int_{\mathbb{W}} e^f d\gamma\right) \leq \mathbb{E}\left(f(B + U) - \frac{1}{2}\|U\|^2\right) + \epsilon,$$

which concludes the proof. \square

2.5 Brownian bridges

A measure μ on \mathbb{W} satisfying

$$\mu(dw) = \rho(w_1) \gamma(dw) \quad (11)$$

where ρ is some density on (\mathbb{R}^d, γ_d) is said to be a Brownian bridge. It can be seen as the law of a Brownian motion conditioned to have law $\rho(x)\gamma_d(dx)$ at time 1.

Lemma 10. *Let ν have density ρ with respect to γ_d , we have*

$$H(\nu \mid \gamma_d) = \inf_{\mu} \left(H(\mu \mid \gamma) \right)$$

where the infimum is on all probability measures satisfying $\mu \circ (x_1)^{-1} = \nu$. The infimum is attained when μ is the bridge (11).

In other words, among all processes having law ν at time 1, the bridge minimizes the relative entropy. This is essentially a particular case of (3), see also [4] and [17, page 161].

Assume that ρ is differentiable and that $\nabla \rho \in L^2(\gamma_d)$. Then $F(w) = \rho(w_1)$ belongs to \mathbb{D}^2 and has Malliavin derivative

$$DF(w) = \nabla \rho(w_1) \mathbf{1}_{[0,1]}.$$

By Lemma 3 the Föllmer process of the bridge μ is such that

$$u_t = E^\mu(\nabla \log(\rho)(w_1) \mid \mathcal{G}_t) \mathbf{1}_{[0,1]}(t), \quad \mu - a.s.$$

We obtain the following result.

Lemma 11. *Under μ , the process $(u_t)_{t \in [0,1]}$ is a martingale. In particular*

$$E^\mu(u_t) = E^\mu \nabla \log(\rho)(w_1) = E^\gamma \nabla \rho(w_1) = \int_{\mathbb{R}^d} x \nu(dx).$$

Now assume that ρ and $\nabla \rho$ are Lipschitz and that $\rho \geq \epsilon$, so that the bridge μ belongs to \mathcal{S} . It is easily seen that u_t can also be written as

$$u_t(w) = \nabla \log P_{1-t} \rho(w_t) \mathbf{1}_{[0,1]}(t),$$

where P_t denotes the heat semigroup on \mathbb{R}^d :

$$\partial_t P_t = \frac{1}{2} \Delta P_t.$$

The stochastic differential equation (7) becomes

$$X_t = B_t + \int_0^{t \wedge 1} \nabla \log(P_{1-s} \rho)(X_s) ds, \quad t \geq 0. \quad (12)$$

By Lemma 6, there is a unique strong solution. Combining Lemma 10 with Theorem 4 we obtain the following dual formulation of Borell's result (1).

Theorem 12. *Let ν and ρ be as above. Then*

$$H(\nu \mid \gamma_d) = \inf_U \left(\frac{1}{2} \mathbb{E} \|U\|^2 \right)$$

where the infimum is taken on all drifts U satisfying $B_1 + U_1 = \nu$ in law. The infimum is attained by the drift

$$U_t = \int_0^{t \wedge 1} \nabla \log(P_{1-s}\rho)(X_s) \, ds,$$

where X is the unique solution of (12).

3 Applications

Following Borell, we now derive functional inequalities from the representation formula. Let us point out that in all but one applications we use Proposition 1 and Theorem 2 rather than Theorem 7.

3.1 Transportation cost inequality

Let T_2 be the transportation cost for the Euclidean distance squared: given two probability measures μ and ν on \mathbb{R}^d

$$T_2(\mu, \nu) = \inf \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, d\pi(x, y) \right)^{1/2}, \quad (13)$$

where the infimum is taken over all couplings π of μ and ν , namely all probability measures on the product space $\mathbb{R}^d \times \mathbb{R}^d$ having marginals μ and ν . There is a huge literature about this optimization problem, usually referred to as Monge-Kantorovitch problem, see Villani's book [25]. Talagrand's inequality asserts that

$$T_2(\nu, \gamma_d) \leq 2 H(\nu \mid \gamma_d)$$

for every probability measure ν on \mathbb{R}^d . The purpose of this subsection is to prove a Wiener space version of this inequality.

On Wiener space the natural definition of T_2 involves the norm of the Cameron-Martin space \mathbb{H} : given two probability measures μ, ν on $(\mathbb{W}, \mathcal{B})$

$$T_2(\mu, \nu) = \inf \left(\int_{\mathbb{W} \times \mathbb{W}} \|w - w'\|^2 \, \pi(dw, dw') \right),$$

where the infimum is taken over all couplings π of μ and ν such that $w - w' \in \mathbb{H}$ for π -almost all (w, w') .

Theorem 13. *Let μ be a probability measure on $(\mathbb{W}, \mathcal{B})$. Then*

$$T_2(\mu, \gamma) \leq 2 H(\mu \mid \gamma).$$

Here is a short proof based of Theorem 2. Fair enough, Feyel and Üstünel [13] have a very similar argument.

Proof. Assume that μ is absolutely continuous with respect to γ (otherwise $H(\mu | \gamma) = +\infty$). According to Theorem 2 there exists a Brownian motion B and a drift U such that $B + U$ has law μ and

$$H(\mu | \gamma) = \frac{1}{2} \mathbb{E} \|U\|^2.$$

Then $(B, B + U)$ is a coupling of (γ, μ) and by definition of T_2

$$T_2(\mu, \gamma)^2 \leq \mathbb{E} \|U\|^2 = 2 H(\mu | \gamma). \quad \square$$

Let us point out that Talagrand's inequality can be recovered easily from this theorem, applying it to a Brownian bridge. Details are left to the reader.

3.2 Logarithmic Sobolev inequality

In this section we prove the logarithmic Sobolev inequality for the Wiener measure, which extends the classical log-Sobolev inequality for the Gaussian measure, due to Gross [18]. When μ is a measure on $(\mathbb{W}, \mathcal{B}, \gamma)$ with density F such that DF is well defined, the Fisher information of μ is

$$I(\mu | \gamma) = \int_{\mathbb{W}} \frac{\|DF\|^2}{F} d\gamma = \int_{\mathbb{W}} \left\| \frac{DF}{F} \right\|^2 d\mu.$$

Theorem 14. *Let μ have density F with respect to γ and assume that $F \in \mathbb{D}^2$. Then*

$$H(\mu | \gamma) \leq \frac{1}{2} I(\mu | \gamma). \quad (14)$$

Proof. We consider the probability space $(\mathbb{W}, \mathcal{B}, \mu)$. Recall that $(\mathcal{G}_t)_{t \geq 0}$ is the filtration of the coordinate process. By Theorem 2 and Lemma 3, letting

$$u_t = \mathbb{E}^\mu \left(\frac{D_t F}{F} \mid \mathcal{G}_t \right)$$

we have

$$H(\mu | \gamma) = \frac{1}{2} \mathbb{E}^\mu \int_0^\infty |u_t|^2 dt.$$

By Jensen's inequality

$$\mathbb{E}^\mu |u_t|^2 \leq \mathbb{E}^\mu \left| \frac{D_t F}{F} \right|^2$$

so that

$$H(\mu | \gamma) \leq \frac{1}{2} \mathbb{E}^\mu \left\| \frac{DF}{F} \right\|^2$$

which is the result. \square

This may not be the most straightforward proof, see [9]. Let us emphasize that applying (14) to a Brownian bridge yields the usual log-Sobolev inequality. More precisely, let ν be a probability measure on \mathbb{R}^d having a smooth density ρ with respect to γ_d and let μ be the measure on \mathbb{W} given by

$$\mu(dw) = \rho(w_1) \gamma(dw).$$

Then $H(\nu \mid \gamma_d) = H(\mu \mid \gamma)$. On the other hand letting $F(w) = \rho(w_1)$ we have

$$DF(w) = \nabla \rho(w_1) \mathbf{1}_{[0,1]},$$

which implies easily that $I(\nu \mid \gamma_d) = I(\mu \mid \gamma)$. Thus (14) becomes

$$H(\nu \mid \gamma_d) \leq \frac{1}{2} I(\nu \mid \gamma_d).$$

3.3 Shannon's inequality

Given a random vector η on \mathbb{R}^d having density ρ with respect to the Lebesgue measure, Shannon's entropy is defined as

$$S(\eta) = - \int_{\mathbb{R}^d} \rho \log(\rho) \, dx.$$

In other words $S(\eta) = -H(\nu \mid \lambda_d)$ where ν is the law of η and λ_d is the Lebesgue measure on \mathbb{R}^d .

Theorem 15. *Let η, ξ be independent random vectors on \mathbb{R}^d and $\theta \in [0, \pi/2]$*

$$S(\cos(\theta)\eta + \sin(\theta)\xi) \geq \cos(\theta)^2 S(\eta) + \sin(\theta)^2 S(\xi). \quad (15)$$

This inequality plays a central role in information theory, see [12] for an overview on the topic.

Proof. Let ν_θ be the law of $\cos(\theta)\eta + \sin(\theta)\xi$. By Theorem 2, Lemma 10 and Lemma 11 there exists a Brownian motion X and a drift U such that

- $X_1 + U_1$ has law ν_0 .
- $H(\nu_0 \mid \gamma_d) = \mathbb{E}\|U\|^2/2$.
- $\mathbb{E}(U) = \mathbb{E}(\eta) \mathbf{1}_{[0,1]}$.

Similarly, there exists a Brownian motion Y and a drift V satisfying the corresponding properties for $\nu_{\pi/2}$. Besides, we can clearly assume that Y is independent of X . Then $\cos(\theta)X + \sin(\theta)Y$ is a Brownian motion and

$$\cos(\theta)X_1 + \sin(\theta)Y_1 + \cos(\theta)U_1 + \sin(\theta)V_1$$

has law ν_θ . By Proposition 1 and Lemma 10

$$H(\nu_\theta \mid \gamma_d) \leq \frac{1}{2} \mathbb{E}\|\cos(\theta)U + \sin(\theta)V\|^2.$$

Denoting the inner product in \mathbb{H} by $\langle \cdot, \cdot \rangle$ we have

$$\mathbb{E}\langle U, V \rangle = \langle \mathbb{E}U, \mathbb{E}V \rangle = (\mathbb{E}\eta) \cdot (\mathbb{E}\xi),$$

so that

$$\begin{aligned} \mathbb{H}(\nu_\theta \mid \gamma_d) &\leq \cos(\theta)^2 \mathbb{H}(\nu_0 \mid \gamma_d) + \sin(\theta)^2 \mathbb{H}(\nu_{\pi/2} \mid \gamma_d) \\ &\quad + \cos(\theta) \sin(\theta) (\mathbb{E}\eta) \cdot (\mathbb{E}\xi). \end{aligned}$$

This is easily seen to be equivalent to (15). \square

3.4 Brascamp-Lieb inequality

Let us focus on a family of inequalities dating back to Brascamp and Lieb's article [8] on optimal constants in Young's inequality. Since then a number of nice alternate proofs have been discovered, see [3, 10] and the survey article [1]. This subsection is inspired by the (unpublished) proof of Maurey relying on Borell's formula.

Let E be a Euclidean space, let E_1, \dots, E_m be subspaces and for all i let P_i be the orthogonal projection with range E_i . The crucial hypothesis is the so-called frame condition: there exist c_1, \dots, c_m in \mathbb{R}_+ such that

$$\sum_{i=1}^m c_i P_i = \text{id}_E. \quad (16)$$

Let $x \in E$, we then have $|x|^2 = (\sum c_i P_i x) \cdot x$ and since P_i is an orthogonal projection

$$|x|^2 = \sum_{i=1}^m c_i |P_i x|^2. \quad (17)$$

From now on \mathbb{W} denotes the space of continuous paths taking values in E and starting from 0 and γ denotes the Wiener measure on \mathbb{W} . The spaces \mathbb{W}_i and measures γ_i are defined similarly.

Theorem 16. *Under the frame condition, for every probability measure μ on \mathbb{W} we have*

$$\mathbb{H}(\mu \mid \gamma) \geq \sum_{i=1}^m c_i \mathbb{H}(\mu_i \mid \gamma_i),$$

where $\mu_i = \mu \circ P_i^{-1}$ is the push-forward of μ by the projection P_i .

Proof. According to Theorem 2 there exists a standard Brownian motion B on E and a drift U such that $B + U$ has law μ and

$$\mathbb{H}(\mu \mid \gamma) = \frac{1}{2} \mathbb{E} \|U\|^2.$$

Since P_i is an orthogonal projection, the process $P_i B$ is a standard Brownian motion on E_i . Also $P_i B + P_i U$ has law $\mu \circ P_i^{-1} = \mu_i$. By Proposition 1

$$\mathbb{H}(\mu_i \mid \gamma_i) \leq \frac{1}{2} \mathbb{E} \|P_i U\|^2, \quad i = 1, \dots, m.$$

On the other hand, the frame condition (17) implies easily that

$$\|U\|^2 = \sum_{i=1}^n c_i \|P_i U\|^2$$

pointwise. Taking expectation yields the result. \square

As observed by Carlen and Cordero [10], this super-additivity property of the relative entropy is equivalent to the following Brascamp-Lieb inequality.

Corollary 17. *Under the frame condition, given m functions $F_i: \mathbb{W}_i \rightarrow \mathbb{R}_+$, we have*

$$\int_{\mathbb{W}} \prod_{i=1}^m (F_i \circ P_i)^{c_i} d\gamma \leq \prod_{i=1}^m \left(\int_{\mathbb{W}_i} F_i d\gamma_i \right)^{c_i}.$$

When the functions F_i depend only on the point w_1 rather than on the whole path w we recover the usual Brascamp-Lieb inequality for the Gaussian measure.

3.5 Reversed Brascamp-Lieb inequality

Again E is a Euclidean space and E_1, \dots, E_m are subspaces satisfying the frame condition (16). Observe that if x_1, \dots, x_m belong to E_1, \dots, E_m respectively, then for any $y \in E$, the Cauchy-Schwarz inequality and (17) yield

$$\begin{aligned} \left(\sum_{i=1}^m c_i x_i \right) \cdot y &= \sum_{i=1}^m c_i (x_i \cdot P_i y) \\ &\leq \left(\sum_{i=1}^m c_i |x_i|^2 \right)^{1/2} \left(\sum_{i=1}^m c_i |P_i y|^2 \right)^{1/2} \\ &= \left(\sum_{i=1}^m c_i |x_i|^2 \right)^{1/2} |y|. \end{aligned}$$

Hence

$$\left| \sum_{i=1}^m c_i x_i \right|^2 \leq \sum_{i=1}^m c_i |x_i|^2. \quad (18)$$

Let \mathcal{S}_i be the class of probability measures on E_i which satisfy the conditions of Definition 5, replacing \mathbb{R}^d by E_i . Here is the reversed version of Theorem 16.

Theorem 18. *Given m probability measures μ_1, \dots, μ_m belonging to $\mathcal{S}_1, \dots, \mathcal{S}_m$ respectively, there exist m processes X_1, \dots, X_m (defined on the same probability space) such that*

1. X_i has law μ_i for all $i = 1, \dots, m$.

2. Letting μ be the law of $\sum c_i X_i$ we have

$$H(\mu \mid \gamma) \leq \sum_{i=1}^m c_i H(\mu_i \mid \gamma_i).$$

Proof. Again let B be a standard Brownian motion on E . For $i = 1, \dots, m$, the process $P_i B$ is a standard Brownian motion on E_i . Since $\mu_i \in \mathcal{S}_i$ there exists a drift U_i such that the process $X_i = P_i B + U_i$ has law μ_i and

$$H(\mu_i \mid \gamma_i) = \frac{1}{2} \mathbb{E} \|U_i\|^2.$$

Let $X = \sum c_i X_i$ and let μ be the law of X . Since $\sum c_i P_i$ is the identity of E

$$X = B + \sum_{i=1}^m c_i U_i.$$

By Proposition 1, we get

$$H(\mu \mid \gamma) \leq \frac{1}{2} \mathbb{E} \left\| \sum_{i=1}^m c_i U_i \right\|^2.$$

On the other hand (18) easily implies that

$$\left\| \sum_{i=1}^m c_i U_i \right\|^2 \leq \sum_{i=1}^m c_i \|U_i\|^2,$$

pointwise. Taking expectation we get the result. \square

This sub-additivity property of the entropy is a multi-marginal version of the *displacement convexity* property put forward by Sturm [22]. By duality, we obtain the following reversed Brascamp-Lieb inequality.

Corollary 19. *Assuming the frame condition, given m functions $F_i: \mathbb{W}_i \rightarrow \mathbb{R}_+$ bounded away from 0, and a function $G: \mathbb{W} \rightarrow \mathbb{R}_+$ satisfying*

$$\prod_{i=1}^m F_i(w_i)^{c_i} \leq G\left(\sum_{i=1}^m c_i w_i\right) \quad (19)$$

for all $(w_1, \dots, w_m) \in \mathbb{W}_1 \times \dots \times \mathbb{W}_m$, we have

$$\prod_{i=1}^m \left(\int_{\mathbb{W}_i} F_i d\gamma_i \right)^{c_i} \leq \int_{\mathbb{W}} G d\gamma.$$

Proof. By Lemma 8, for every i , there exists a measure $\mu_i \in \mathcal{S}_i$ such that

$$\log \left(\int_{\mathbb{W}_i} F_i d\gamma_i \right) \leq \int_{\mathbb{W}_i} \log(F_i) d\mu_i - H(\mu_i \mid \gamma_i) + \epsilon.$$

Let X_1, \dots, X_m be the random processes given by the previous theorem, let $X = \sum c_i X_i$ and let μ be the law of X . Then by duality and the hypothesis (19) we get

$$\begin{aligned} \log\left(\int_{\mathbb{W}} G \, d\gamma\right) &\geq \mathbb{E} \log(G)(X) - H(\mu \mid \gamma) \\ &\geq \mathbb{E}\left(\sum_{i=1}^m c_i \log(F_i)(X_i)\right) - H(\mu \mid \gamma). \end{aligned}$$

Since $H(\mu \mid \gamma) \leq \sum c_i H(\mu_i \mid \gamma_i)$, this is at least

$$\sum c_i \left(\log\left(\int_{\mathbb{W}_i} F_i \, d\gamma_i\right) - \epsilon \right).$$

Letting ϵ tend to 0 yields the result. \square

Again when the functions depend only on the value of the path at time 1, we recover the reversed Brascamp-Lieb inequality for the Gaussian measure, which is due to Barthe [2].

Acknowledgements

The author is grateful to Patrick Cattiaux and Massimiliano Gubinelli for communicating references and to Christian Léonard, Bernard Maurey and Patrick Cattiaux again for valuable discussions.

References

- [1] K. Ball. Convex geometry and functional analysis, in *Handbook of the Geometry of Banach Spaces*, vol. 1, edited by W.B. Johnson and J. Lindenstrauss, North Holland (2001) 161–194.
- [2] F. Barthe. On a reverse form of the Brascamp-Lieb inequality, *Invent. Math.* 134 (2) (1998) 335–361.
- [3] F. Barthe and N. Huet. On Gaussian Brunn-Minkowski inequalities, *Studia Math.* 191 (3) (2009) 283–304.
- [4] F. Baudoin. Conditioned stochastic differential equations: theory, examples and application to finance. *Stochastic Process. Appl.* 100 (2002) 109–145.
- [5] C. Borell. Diffusion equations and geometric inequalities, *Potential Anal.* 12 (1) (2000) 49–71.
- [6] C. Borell. Isoperimetry, log-concavity, and elasticity of option prices, in *New Directions in Mathematical Finance*, edited by P. Wilmott and H. Rasmussen, Wiley (2002) 73–91.

- [7] M. Boué and P. Dupuis. A variational representation for certain functionals of Brownian motion, *Ann. Probab.* 26 (4) (1998) 1641–1659.
- [8] H. J. Brascamp and E. H. Lieb. Best constants in Youngs inequality, its converse and its generalization to more than three functions. *Adv. Math.* 20 (1976) 151–173.
- [9] M. Capitaine, E.P. Hsu and M. Ledoux. Martingale representation and a simple proof of logarithmic Sobolev inequalities on path spaces, *Electron. Comm. Probab.* 2 (1997) 71–81.
- [10] E. Carlen and D. Cordero-Erausquin. Subadditivity of the entropy and its relation to Brascamp-Lieb type inequalities, *Geom. Funct. Anal.* 19 (2009) 373–405.
- [11] D. Cordero-Erausquin and M. Ledoux. The geometry of Euclidean convolution inequalities and entropy, *Proc. Amer. Math. Soc.* 138 (8) (2010) 2755–2769.
- [12] A. Dembo, T.M. Cover and J.A. Thomas. Information Theoretic Inequalities, *IEEE Trans. Inform. Theory* 37 (6) (1991) 1501–1518.
- [13] D. Feyel and A.S. Üstünel, Measure transport on Wiener space and the Girsanov theorem, *C. R. Math. Acad. Sci. Paris* 334 (11) (2002) 1025–1028.
- [14] W.H. Fleming and H.M. Soner. Controlled Markov processes and viscosity solutions. Second edition. *Stochastic Modelling and Applied Probability*, 25. Springer, New York, 2006.
- [15] H. Föllmer, An entropy approach to the time reversal of diffusion processes, in *Stochastic differential systems (Marseille-Luminy, 1984)*, Lecture Notes in Control and Inform. Sci. 69, Springer (1985) 156–163.
- [16] H. Föllmer. Time reversal on Wiener space, in *Stochastic processes—mathematics and physics (Bielefeld, 1984)*, Lecture Notes in Math. 1158, Springer (1986) 119–129.
- [17] H. Föllmer. Random fields and diffusion processes. in *École d’Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, Lecture Notes in Math. 1362, Springer (1988) 101–203.
- [18] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.* 97 (4) (1975) 1061–1083.
- [19] R.S. Liptser and A.N. Shiryaev. Statistics of random processes. Vol. 1. General theory. Translated from the Russian original by A.B. Aries. Second edition. *Applications of Mathematics* 5. Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2001.

- [20] D. Nualart. The Malliavin calculus and related topics. Second edition. Probability and its Applications. Springer-Verlag, Berlin, 2006.
- [21] L.C.G. Rogers and D. Williams. Diffusions, Markov processes, and martingales. Vol. 2. Itô calculus. Reprint of the second (1994) edition. Cambridge Mathematical Library. Cambridge University Press, 2000.
- [22] K.T. Sturm, On the geometry of metric measure spaces. I. Acta Math. 196 (1) (2006) 65–131.
- [23] M. Talagrand. Transportation cost for Gaussian and other product measures, Geom. Funct. Anal. 6 (3) (1996) 587–600.
- [24] S.R.S Varadhan. Large deviations and applications. CBMS-NSF Regional Conference Series in Applied Mathematics, 46. SIAM, Philadelphia, 1984.
- [25] C. Villani. Optimal transport. Old and new. Grundlehren der Mathematischen Wissenschaften, 338. Springer-Verlag, Berlin, 2009.